Statistical Considerations for the Analysis of Ordinal Outcomes in Randomized Controlled Trials

by

So Yung Choi

A thesis submitted to The Johns Hopkins University
in conformity with the requirements for the degree of
Master of Science

Baltimore, Maryland
April, 2016

Abstract

In randomized controlled trials, ordinal outcomes are commonly used as primary endpoints to measure the efficacy of interventions. Conventionally, the ordinal outcome has been analyzed on the original scale, by first creating a binary measure or by treating the outcome as a continuous variable. It has been suggested that analyzing the data on the original ordinal scale provides the most powerful test for detecting a treatment effect. However, there are situations when dichotomizing the ordinal outcome yields higher power.

In this thesis, we review the conventionally used statistical methods for analyzing ordinal outcomes, and apply these methods to simulated hypothetical trials. The simulated hypothetical trials are defined based on different distributional assumptions for the control arm. To test for a treatment effect, we apply the cumulative logit model to the ordinal outcome and the Fisher's exact test to the dichotomized outcome. The power to detect a treatment effect is compared across these two methods for each control arm distribution and a variety of treatment effect sizes.

The power to detect the treatment effect depends on the control arm distribution and the anticipated treatment effect. We found that dichotomizing the ordinal outcome can yield higher power to detect a treatment effect if the difference in the proportion of patients at a single level of the ordinal outcome is large and the remaining differences are spread over the remaining categories as opposed to shifted into a single category.

Acknowledgements

I am deeply indebted to my advisor Dr. Elizabeth Colantuoni for her continuous support throughout my time in this program from the first day of school to now with the submission of this thesis. I feel extremely grateful to Dr. Colantuoni for her patience, enthusiasm, and kindness. Without her counsel and encouragement, I would not have been able to finish my thesis.

My sincere appreciation also goes to Dr. Elizabeth Sugar and Dr. Gayane Yenokyan, who have been co-advising this thesis project with their immense expertise, experience and insightful guidance.

I would also like to thank all the people in the department. I enjoyed my time here, and truly appreciate the friendly environment and support created by the amazing staff, faculty, and students.

And lastly I want to express my utmost love and gratitude to my family, for all the love and support during my life and helping me grow up to become a happy person.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Ordinal outcomes are commonly used as primary endpoints in randomized controlled trials; however, defining the treatment effect when the primary endpoint is an ordinal outcome is challenging. Defining the treatment effect requires careful considerations; including the interpretation of the ordered categories, the clinical relevance of each ordered category, the distribution of the outcome within the control arm and the anticipated effect of the treatment on the outcome.

It is common to apply a fixed dichotomy to an ordinal outcome to provide a simplified definition of the treatment effect. However, several recent papers have suggested that preserving the characteristics of the ordinal outcome (natural ordering and number of ordered levels) in the definition of and statistical analysis estimating the treatment effect provides higher power relative to the fixed dichotomy [1, 2].

To address this conflict, in this thesis, we review the conventionally used treatment effect definitions and subsequent statistical methods when analyzing an ordinal outcome in a randomized clinical trial. We apply these methods to data from two completed Phase II trials on stroke patients, the MISTIE II trial and the CLEAR II

trial. Using the MISTIE II and CLEAR II trial data, we simulate hypothetical trials and demonstrate the power to detect the treatment effect of an ordinal outcome vs. fixed dichotomy. We conclude by providing guidance for researchers to optimally design and analyze a randomized clinical trial with an ordinal outcome.

1.1 Definitions and notation

Throughout this thesis, we consider a 1:1 randomized controlled trial, such that n subjects are randomly assigned to receive treatment $A \in \{0, 1\}$, where $A \sim Bernoulli(0.5)$. The outcome $Y \in \{1, \cdots, K\}$ represents an ordered categorical outcome with K possible values. Define $p_k = Pr(Y = k)$ for $k = 1, \cdots, K$, with $\sum_{k=1}^{K} p_k = 1$. The data consists of (A_i, Y_i) for each subject i in $i = 1, \cdots, n$, and we define p_{0k} and p_{1k} as $Pr(Y_i = k \mid A_i = 0)$ and $Pr(Y_i = k \mid A_i = 1)$, respectively. The outcome Y with K ordinal levels and the treatment assignment can be displayed in a 2 x K contingency table (Table 1.1) with n_{0k} and n_{1k} denoting the number of treated and control patients in the $Y_i = k$ category, respectively. We additionally define $n_{.k}$ as the number of patients in the $Y = k$ outcome category regardless of treatment assignment, i.e. $n_{.k} = n_{0k} + n_{1k}$. And let $n_{0.} = \sum_{k=1}^{K} n_{0k}$ and $n_{1.} = \sum_{k=1}^{K} n_{1k}$ be the number of patients in the control group ($A = 0$) and in the treated group ($A = 1$), respectively.

Table 1.1: 2 x K contingency table

	$Y = 1$	$Y = 2$	$\cdots$	$Y = K$	
$A = 0$	n_{01}	n_{02}	$\cdots$	n_{0K}	$n_{0.}$
$A = 1$	n_{11}	n_{12}	$\cdots$	n_{1K}	$n_{1.}$
	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.K}$	n

Chapter 2

Review of statistical methods

Researchers have historically defined the treatment effect and appropriate statistical analysis for an ordinal outcome using one of three approaches: 1) maintain the outcome on the original scale, 2) dichotomize the ordinal outcome to create a binary outcome, and 3) treat the ordinal outcome as a continuous variable. In each of the three cases, the treatment effect has a different definition and the corresponding statistical methods to estimate and test the treatment effect vary. In this section, we will review each approach providing the definition of the treatment effect and describing the statistical procedure to estimate and/or test the treatment effect. Table 2.1 summarizes the three approaches by providing a description of summary statistics, null and alternative hypotheses and corresponding statistical methods.

2.1 Ordinal analysis

In the ordinal anlaysis, the outcome variable is treated as the original ordered categorical variable. There are several commonly used statistical methods to define and test for a treatment effect of the ordered outcomes. Some of which utilizes

Table 2.1: Summarized methods based on different approaches of treating ordered categorical outcomes

Outcome is treated as	Summary Statistics	Null Hypothesis	Alternative Hypothesis	Methods
Ordinal	Median and interquartile range	$p_{0k} = p_{1k}$ $for\ all\ k\ = 1, \cdots, K$	$p_{0k} \neq p_{1k}$ $for\ some\ k\ = 1, \cdots, K$	Mann-Whitney U/Wilcoxon test Robust ranks test Cumulative logit model
Dichotomous	Proportion of $Y \leq$ cut-off	$p_{0c} = p_{1c}$ $for\ the\ cut\ point\ c$	$p_{0c} \neq p_{1c}$ $for\ the\ cut\ point\ c$	Fisher's exact test
Continuous	Mean and standard deviation	$\mu_0 = \mu_1$	$\mu_0 \neq \mu_1$	Two-sample t-test

the ranks of the outcomes to test a treatment effect, such as the Mann-Whitney U/Wilcoxon test and robust ranks test. And some takes a parametric approach, such as the cumulative logit model. The goal of each statistical method is to test the null hypothesis that the distribution of the ordinal outcome is the same in the control and treatment group, i.e. $p_{0k} = p_{1k}$ for all k in $1, \cdots, K$, vs $p_{0k} \neq p_{1k}$ for at least one k. The specific ways of stating the null and alternative hypothesis and the statistical approaches for comparing the distributions across the control and treatment groups vary.

2.1.1 Mann-Whitney U/Wilcoxon test

The Mann-Whitney U/Wilcoxon test is used to test the null hypothesis that the outcomes in each treatment group are sampled from the same distribution. To test if the observed samples from the two treatment groups share the same distribution, Mann-Whitney U/Wilcoxon test states its null hypothesis as $Pr(Y_{0j} > Y_{1m}) = Pr(Y_{0j} < Y_{1m}) = \frac{1}{2}$, where Y_{0j} and Y_{1m} are randomly selected observations from the

control and treatment group, respectively. The alternative two-sided hypothesis is $Pr(Y_{0j} > Y_{1m}) \neq Pr(Y_{0j} < Y_{1m}) \neq \frac{1}{2}$ [3].

The Mann-Whitney U/Wilcoxon test ranks all of the observations from lowest to highest, regardless of the treatment assignment. Then the sums of ranks for each treatment group is computed as $S_0 = \sum_{i=1}^{n}(1 - A_i)R_i$ and $S_1 = \sum_{i=1}^{n} A_i R_i$ for the control and treatment group, respectively, where R_i represents the rank of outcome Y_i without differentiating the treatment assignment [4]. Then, the U statistic [5] is the difference in the rank sums between the treatment groups, which can take either value of the U_0 and U_1 defined below.

$$U_0 = n_{0.}n_{1.} + \frac{n_{0.}(n_{0.} + 1)}{2} - S_0$$

$$U_1 = n_{0.}n_{1.} + \frac{n_{1.}(n_{1.} + 1)}{2} - S_1$$

If the U statistic is larger than its critical value, there is statistical evidence to support differences in the distributions of Y between the treatment groups.

In the case of larger sample sizes (typically >10 in each group), we can use a normal approximation for the distribution of U such that

$$U \sim Normal(\frac{n_{0.}n_{1.}}{2}, \frac{n_{0.}n_{1.}(n + 1)}{12})$$

Thus we can calculate the z statistic, $z = \frac{U - \frac{n_{0.}n_{1.}}{2}}{\sqrt{\frac{n_{0.}n_{1.}(n+1)}{12}}}$, and conduct the significance test.

When there are many tied ranks present,such as the case with an ordinal outcome, more complicated corrections are required in order to use the normal approximation. The tied observations are assigned with the average of the ranks when no ties were

assumed. If the ties occur within the same treatment arm, we do not need to make a correction because the rank sums (S_0 and S_1) are not affected by the ties. However when the ties occur across both treatment groups, it affects the variability of each set of ranks, thus the corrected z statistic is now:

$$z = \frac{U - \frac{n_{0.}n_{1.}}{2}}{\sqrt{[\frac{n_{0.}n_{1.}}{n(n-1)}][\frac{(n^3-n)-\sum_1^g(t_m^3-t_m)}{12}]}}$$

where g is the number of sets of tied ranks, and t_m is the number of tied ranks in the m th set of tied ranks[3].

2.1.2 Robust ranks test

The robust ranks test loosens the null hypothesis of Mann-Whitney U/Wilcoxon test, and compares the medians of the ordinal outcomes between the two treatment group, without assuming the distributional equivalence between the two groups. The null hypothesis of the robust ranks test states that $\theta_0 = \theta_1$, where the θ_0 and θ_1 represent the medians of the distributions of the populations for the control and treatment groups, respectively. Correspondingly, the alternative hypothesis of a two-tailed test is stated as $\theta_0 \neq \theta_1$. In the robust ranks test, we order Y_i from smallest to largest value, regardless of treatment assignment, and define R_i as we did for the Mann-Whitney U/Wilcoxon Test. With R_i, we define T_j as the j th observation in the treated group, and C_l as the the l th observation in the control group.

Then for each T_j, let $W(CT_j)$ be the number of observations in the control group that are smaller than T_j. The mean of $W(CT_j)$ can be calculated as $W(CT) = \sum_{j=1}^{n_{1.}} \frac{W(CT_j)}{n_{1.}}$. Similarly, the mean of $W(TC_l)$ can be obtained by $W(TC) = \sum_{l=1}^{n_{0.}} \frac{W(TC_l)}{n_{0.}}$. We define the variances for $W(CT_j)$ and $W(TC_l)$ as $V_T = \sum_{j=1}^{n_{1.}}[W(CT_j - W(CT)]^2$,

$V_C = \sum_{l=1}^{n_{0.}}[W(TC_l - W(TC)]^2$, respectively [3]. We define the test statistic of the robust ranks test, $\grave{W}, as$:

$$\grave{W} = \frac{n_{1.}W(CT) - n_{0.}W(TC)}{2\sqrt{V_T + V_C + W(CT)W(TC)}}$$

If the p-value of the $\grave{W}$ statistic is smaller than its critical value, the null hypothesis is rejected, thus supporting the alternative hypothesis of different medians in the distribution of outcomes in the treatment groups. When the sample sizes are sufficiently large, the distribution of $\hat{W}$ is approximated by the standard normal distribution, thus we can use $\hat{W}$ as an approximate of the z statistic. Similar to the Mann-Whitney U/Wilcoxon test, many ties in the ranks can occur when we have ordered categorical outcomes. In such cases, adjustments are made to account for the ties by redefining $W(CT_j)$ $(W(TC_l))$ as the sum of the number of observations in the control (treated) arm that are less than T_j (C_l) and the half the number of observations in the control (treated) arm that are equal to T_j (C_l).

2.1.3 Cumulative logit model

Unlike the Mann-Whitney U test and the robust rank test, the cumulative logit model defines a parametric model describing differences in the distribution of the ordinal outcome across the treatment groups. When there are K ordered categories for the outcome Y, $K-1$ cumulative logits are defined, each comparing the odds of the response being less than or equal to k across the two treatment groups:

$$logit\left[Pr(Y \leq k \mid A)\right] = log\left[\frac{Pr(Y \leq k \mid A)}{Pr(Y > k \mid A)}\right] = \alpha_k + \beta_k \times A, \text{ for } k = 1, ...K-1$$

α_k represents the log odds of Y being less than or equal to k in the control group ($A = 0$), and β_k represents the difference in the log odds of Y being less than or equal to k comparing the treatment ($A = 1$) and control ($A = 0$) groups.

The cumulative logit model provides a 1-to-1 mapping from the cumulative logits to the distribution of Y for each treatment group $A \in \{0, 1\}$:

$$\alpha_k = log\left[\frac{\sum_k p_{0k}}{1 - \sum_k p_{0k}}\right]$$

$$\alpha_k + \beta_k = log\left[\frac{\sum_k p_{1k}}{1 - \sum_k p_{1k}}\right]$$

In this model, the null hypothesis of no treatment effect reduces to testing if $\beta_k = 0$ for all $k = 1, ..., K - 1$. The alternative hypothesis is $\beta_k \neq 0$ for at least one $k = 1, ..., K - 1$. To test the null hypothesis, a $K - 1$ degree of freedom likelihood ratio test is conducted.

One draw back to this approach is that the treatment effect is defined by $K - 1$ values of β_k and thus is not represented by a single parameter within the regression model. A special case of the cumulative logit model is the proportional odds model, which assumes $\beta_k = \beta$ for all $k = 1, ..., K - 1$ thus defining the treatment effect with a single parameter.

$$log\left[\frac{Pr(Y \leq k \mid A)}{1 - Pr(Y \leq k \mid A)}\right] = \alpha_k + \beta \times A, \text{ for } k = 1, ...K - 1$$

β is the difference in the log odds of Y being less than or equal to k comparing the treatment and control groups. The proportional odds assumption can be often violated in real data. Although there are statistical tests to perform to assess the validity of the proportional odds assumption, such as Brant test [6] and likelihood

ratio test with the cumulative logit model, these tests tend to be underpowered to reject the assumption of proportional odds. In cases where the cumulative logit model is the true model and the proportion odds model is fit, β may or may not be an appropriate summary of the treatment effect within a randomized trial, even when the results of the proportional odds assumption test tells that the assumption is valid.

2.2 Dichotomous analysis

Oridnal outcomes can be dichotomized at a specific threshold to create a binary outcome (i.e. "good" vs. "bad" outcome). We define the dichotomized outcome as $Y_i^* = I(Y_i \leq c)$ for some cut-off $c = 1, ..., K-1$, and $p_{0c} = Pr(Y_i^* = 1|A_i = 0)$ and $p_{1c} = Pr(Y_i^* = 1|A_i = 1)$ for the control and treatment groups, respectively. The null hypothesis for the fixed dichotomy approach is $p_{0c} = p_{1c}$, and the two-sided alternative hypothesis is $p_{0c} \neq p_{1c}$. Several statistical methods can be used to analyze binary outcomes, including logistic regression, the χ^2 test, and Fisher's exact test. In the following section, we provide details of the Fisher's exact test.

Table 2.2: 2 x 2 contingency table

	$Y \leq c$	$Y > c$	
$A = 0$	n_{0c}	$n_{0.} - n_{0c}$	$n_{0.}$
$A = 1$	n_{1c}	$n_{1.} - n_{1c}$	$n_{1.}$
	$n_{.c}$	$n - n_{.c}$	n

2.2.1 Fisher's exact test

When the sample size is small or one of the expected values in a 2 x 2 contingency table is less than 1, the Fisher's exact test can be used to test the independence between a binary outcome and the treatment assignment, see Table 2.2. To reject the null hypothesis that $p_{0c} = p_{1c}$, the observed counts in Table 2.2 must be rare if the data was actually generated under a model where the outcome and treatment are independent. The p-value for the Fisher's exact test is obtained by the following procedures. First, calculate the observed probability, p_{obs} based on the observed 2 x 2 contingency table.

$$p_{obs} = \frac{n_{.c}!(n - n_{.c})!n_{0.}!n_{1.}!}{n!n_{0c}!(n_{0.} - n_{0c})!n_{1c}!(n_{1.} - n_{1c})!}$$

Then, calculate all the other probabilities from the possible 2 x 2 contingency tables with the same marginal frequencies $(n_{.c}, n - n_{.c}, n_{0.}, n_{1.})$ as the observed 2 x 2 contingency table. Finally, the two-sided p-value is obtained by adding up all the probabilities that are less that or equal to p_{obs} [7].

2.3 Continuous analysis

The K ordered categories can be treated as a numeric or continuous variable, and subsequently, statistical methods for comparing two independent population means can be applied such as the two-sample t-test. The null hypothesis of $\mu_0 = \mu_1$ vs. the alternative that $\mu_0 \neq \mu_1$ is tested in the two-sample t-test, where μ_0 and μ_1 are the population means of the K ordered categories which take the numeric values 1 through K. The two-sample t-test assumes that the sample means from the two

treatment groups are normally distributed; this assumption should be valid in studies of relatively large sizes e.g. in phase III trials. The test statistic is computed as:

$$t = \frac{\bar{y_0} - \bar{y_1}}{\frac{(n_{0.}-1)s_0^2+(n_{1.}-1)s_1^2}{n_{0.}+n_{1.}-2} \times \sqrt{\frac{1}{n_{0.}} + \frac{1}{n_{1.}}}}$$

The test statistic t follows a t distribution with $n-2$ degrees of freedom [1]. If the p-value is less than the significance level, we reject the null hypothesis of equal means from the distribution of the outcomes for the two treatment groups.

Chapter 3

Illustrative examples

The methods discussed in the previous section will be applied to two completed phase II randomized clinical trials, MISTIE II and CLEAR II. The focus will be on the hypothesis test of no treatment effect within each trial using the cumulative logit model and the special case where proportional odds are assumed. In addition, we apply a fixed dichotomy approach to each trial using the cut-point defined according to the trials' protocol and test for a treatment effect using Fisher's exact test.

3.1 MISTIE II

Minimally invasive Surgery and rtPA for Intracerebral Hemorrhage Evacuation, the MISTIE II trial is a phase-II randomized clinical trial, completed in 2005 with 96 participants with intracerebral hemorrhage (ICH) recruited from the U.S., Canada, Germany, and U.K. [8]. The primary purpose of the study was to determine the safety and the efficacy of using a combination of minimally invasive surgery (MIS) and clot lysis with recombinant tissue plasminogen activator (rt-PA) to remove ICH. Among the 96 participants, 54 subjects were randomized to receive the combination

of MIS and rt-PA, and 42 subjects were randomized to receive the standard of care medical management as per American Heart Association (AHA) guidelines. Two and four patients from the treatment arm and the control arm were lost to follow-up within the 180-day of the randomization, leaving 52 and 38 treatment and control patients, respectively.

The primary safety outcomes included the 30-day mortality, 7-day procedure-related mortality, 30-day rate of cerebritis, meningitis, bacterial ventriculitis, and 72-hour rate of symptomatic rebleeding. The primary and secondary efficacy outcomes were defined based on the modified Rankin Scale (mRS). The mRS consists of 7 ordered categories (0 to 6) representing the degree of disability of the stroke patient with smaller values indicating less disability: 0 being no symptoms, 1 being no significant disability despite symptoms, 2 being slight disability, 3 being moderate disability, 4 being moderately severe disability, 5 being severe disability, and 6 being dead [9]. The primary efficacy outcome was a binary indicator for the 180-day mRS being at most 3 and the secondary efficacy outcome was the ordinal mRS at 180-days.

Figure 3.1 (a) displays the cumulative distribution function for the 180-day mRS for the MISTIE II trial data among 90 patients who completed the 180-day follow-up. For the primary efficacy endpoint based on the fixed dichotomy with cut-point 3 for the 180-day mRS, the estimated proportion of patients with a successful outcome at 180 days is 0.35 (18/52) and 0.24 (9/38) for the treatment vs. control groups, respectively. There was not a statistically significant difference in the proportion of patients with a successful outcome at 180-days in this phase II trial ($p = 0.18$, Table 3.1) based on the Fisher's exact test. The secondary efficacy analysis compared

the distribution of the 180-day mRS across the two treatment groups. We fit both the cumulative logit model and the proportion odds models to the MISTIE II data, while considering the mRS $\in \{0, 1, 2\}$ as a single category due to no or small observations in the lower categories, and in both analyses we found no statistically significant treatment effect (p = 0.84 for the cumulative logit model and p = 0.4765 for the proportional odds model). From the cumulative logit model, the estimated log odds of better than "slight disability" ($mRS \leq 2$) comparing the treatment and control groups is 0.28 (95% Confidence Interval (CI): -1.03, 1.59), the estimated relative log odds of better than "moderate disability" ($mRS \leq 3$) is 0.53 (95% CI: -0.41, 1.48), the estimated relative log odds of better than "moderately severe disability" ($mRS \leq 4$) is 0.18 (95% CI: -0.67, 1.02), and the estimated relative log odds of better than "severe disability" ($mRS \leq 5$) is 0.10 (95% CI: -0.83, 1.03). Assuming proportional odds across the $K - 1$ ($K = 5$) levels of the ordinal outcomes, the estimated relative log odds of less disability is 0.27 comparing the treatment and control groups (95% CI: -0.48, 1.02). The proportional odds assumption is assessed based on the likelihood ratio test between the cumulative logit model and proportional odds model, and did not have enough evidence to reject the proportional odds assumption (p = 0.82).

3.2 CLEAR II

Clinical Trial on Treatment of Intraventricular Hemorrhage, the CLEAR II trial is a phase-II trial completed in 2008 [10]. The purpose of this trial was to determine the safety of and the lowest efficacious dose and dose frequency of rt-PA injected via intraventricular catheter (IVC) for the treatment of patients with small ICH

(ICH$\leq$30) and large intraventricular hemorrhage (IVH) who already have an IVC placed. There were 3 stages in the CLEAR II trial: CLEAR Safety study, CLEAR A (dose finding trial), and CLEAR B (dose escalation trial). A total of 100 patients were enrolled throughout the 3 stages of the trial, recruited from the U.S., Canada, Germany, and U.K., and of 100 participants, 78 patients were allocated to receive different doses and dose frequencies of the rt-PA treatment, and 22 patients were randomized to receive placebo of normal saline during the stage of CLEAR safety study, in which both the intervention and placebo drug was injected via the IVC.

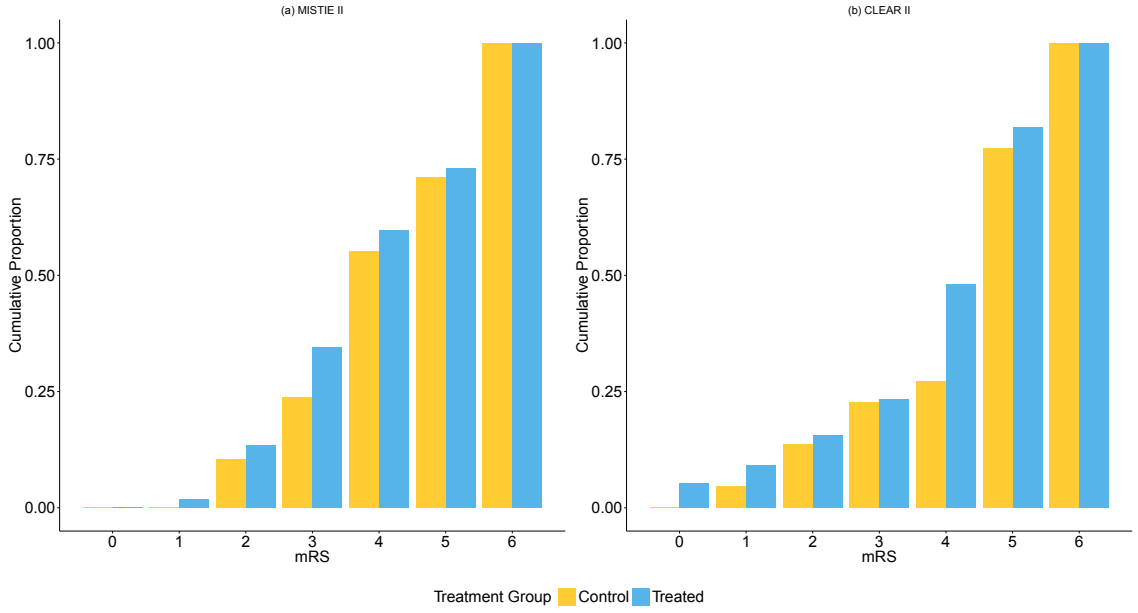
The primary outcomes were three safety outcomes within the 30-day follow-up: mortality, incidence of bacterial ventriculitis, and rate of symptomatic bleeding events. In our reanalysis of this phase II study, we will focus on a efficacy outcome, the 30-day mRS. The cumulative distribution function for the CLEAR II trial data is presented in Figure 3.1 (b) based on the 30-day mRS from the 99 patients in total (22 in the control arm and 77 in the treated arm due to 1 person lost to follow-up). When the fixed dichotomy approach was used with a cut-point of 4 for the 30-day mRS, the estimated proportion of patients with a successful outcome ($Y \leq 4$) at 30-days is 0.48 (37/77) and 0.27 (6/22) for the treatment and control groups, respectively. There was no statistically significant treatment effect in this phase II trial (p = 0.09, Table 3.1) based on the Fisher's exact test. When the distributions of the 30-day mRS were compared across the two treatment groups, treating the mRS as the original ordinal scale, with combining the mRS $\in \{0, 1\}$ into a single category due to no observation in mRS of 0 among the control patients, there was no evidence of a treatment effect (p = 0.25 based on the cumulative logit model and

p = 0.24 based on the proportional odds model). From the cumulative logit model, the estimated log odds of better than "no significant disability despite symptoms" ($mRS \leq 1$) comparing the treatment and control groups is 0.74 (95% CI: -1.41, 2.89), the estimated log odds ratio of better than "slight disability" ($mRS \leq 2$) is 0.16 (95% CI: -1.21, 1.52), the estimated relative log odds of better than "moderate disability" ($mRS \leq 3$) is 0.04 (95% CI: -1.09, 1.16), the estimated relative log odds of better than "moderately severe disability" ($mRS \leq 4$) is 0.90 (95% CI: -0.14, 1.94), and the estimated relative log odds of better than "severe disability" ($mRS \leq 5$) is 0.28 (95% CI: -0.87, 1.43). When proportional odds across the $K-1$ ($K = 6$) levels of the ordinal outcomes are assumed, the estimated relative odds of less disability is 0.52 (95% CI: -0.35, 1.38). Also, the assumption of proportional odds is assessed using likelihood ratio test with the cumulative logit model, and did not show enough evidence to reject the proportional odds assumption (p = 0.27).

Table 3.1: The results of significant test (p-values) based on the MISTIE II and CLEAR II trial data using different statistical methods

Methods	MISTIE II	CLEAR II
Fisher's Exact Test	0.1772	0.0940
Cumulative Logit Model	0.8397	0.2516
Proportional Odds Model	0.4765	0.2388

Figure 3.1: Cumulative distributions of mRS by the treatment arms for the MISTIE II and CEAR II trials



Chapter 4

Simulation study

4.1 Simulation design

We performed a simulation study to understand the characteristics of the distributions of the ordinal outcome in the treatment and control group such that the use of the cumulative logit model, the proportional odds model or the fixed dichotomy are optimal in the sense of achieving high statistical power to detect a treatment effect. Specifically, we sought to identify characteristics of the distributions of the ordinal outcome where the fixed dichotomy yields higher power than the cumulative logit model, and vice versa. In addition, when it is reasonable to assume proportional odds, is there a benefit in terms of power to detect a treatment effect to chosing the fixed dichotomy relative to the proportional odds model.

The data for the simulation was generated as follows: $n = 500$ subjects were allocated 1:1 to receive treatment or control (250 each per treatment group), the outcome for subject is a value from a K level ordinal outcome Y, according to the following cumulative logit model:

$$log\left[\frac{Pr(Y \leq k \mid A)}{Pr(Y > k \mid A)}\right] = \alpha_k + \beta_k \times A, \text{ for } k = i, ...K-1$$

Note that this model reduces to the proportional odds model when β_k are the same for all k.

Our simulation study mimics features of the MISTIE II and CLEAR II trial in defining the outcome Y as the mRS, where for simplicity we let $K = 3$, such that Y is defined as:

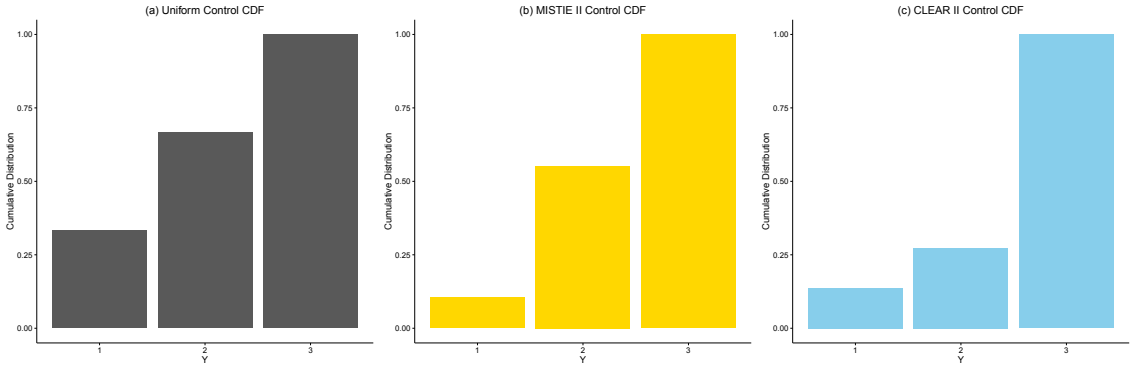
$$Y = \begin{cases} 1 & \text{if } mRS \in \{0, 1, 2\} \\ 2 & \text{if } mRS \in \{3, 4\} \\ 3 & \text{if } mRS \in \{5, 6\} \end{cases}$$

Then α_1 represents the log odds of $Y = 1$ among patients in the control group, and α_2 represents the log odds of $Y \leq 2$ among patients in the control group. β_1 is the log odds ratio of $Y = 1$ comparing the treatment and control groups, and β_2 is the the log odds ratio of $Y \leq 2$ comparing the treatment and control groups.

For our simulation study, three different distributions for the ordinal outcome Y within the control group are considered: the uniform distribution, and the control distributions observed in the MISTIE II and CLEAR II trials. The graphical display of theses three control group distributions is presented in Figure 4.1.

We generated the data from the cumulative logit model. The $\{\alpha_k\}$ were uniquely defined for each control group distribution, such that for the uniform control group distribution, $(\alpha_1 = -0.69, \alpha_2 = 0.69)$,$(\alpha_1 = -2.14, \alpha_2 = 0.21)$ for the MISTIE II control group distribution, and $(\alpha_1 = -1.84, \alpha_2 = -0.97)$ for the CLEAR II control group distribution. We allowed the effect of treatment on the ordinal outcome to be in a wide range, β_1 and β_2 each ranging from -0.80 to 0.80.

Figure 4.1: Graphical display of the three control group distributions



For each control group distribution (i.e. (α_1, α_2)) and (β_1, β_2) pair, we generated 10,000 simulated studies with the sample size of 250 subjects per group. Within each simulated study, we computed the following: i) the test of treatment effect based on the cumulative logit model and ii) the test of treatment effect assuming a proportional odds model. The power to detect a treatment effect was computed by taking the proportion of tests that rejected the null hypothesis across the 10,000 simulated studies. In addition, we used exact binomial calculations to compute the power of the fixed dichotomy based on cut-points $Y = 1$ and $Y = 2$. We display the power to detect a treatment effect within the cumulative logit model, the proportional odds model and the fixed dichotomy as a function of two odds ratios (e^{β_1} and e^{β_2}) using a heat map for each assumed control arm distribution.

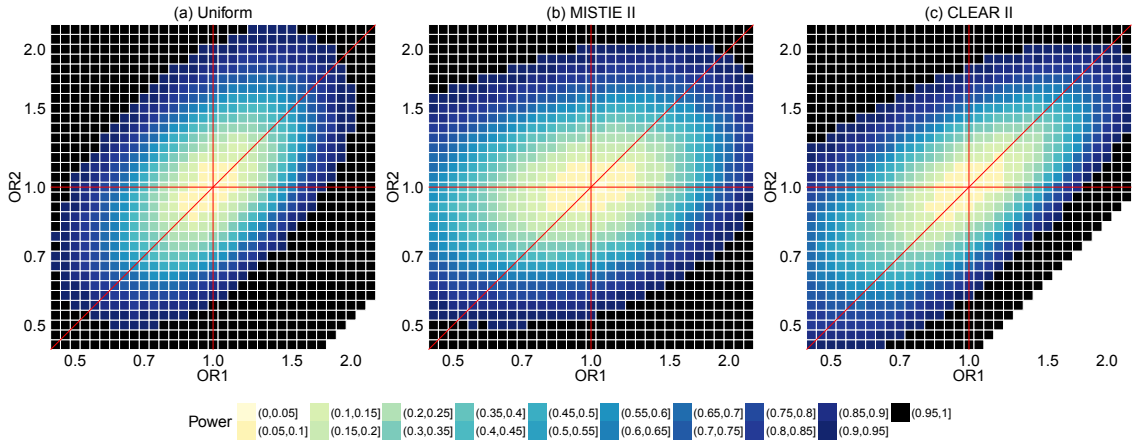
For the uniform and CLEAR II control group distributions, a true discrete probability distribution could not be defined for the treatment group when β_1 is small and β_2 is large. Therefore, there are several combinations of (β_1, β_2) for which power is not evaluated.

4.2 Results

4.2.1 Cumulative logit model

Figure 4.2 displays the power to detect a treatment effect based on the cumulative logit model as a function of OR_1 and OR_2 ($OR_1 = e^{\beta_1}$ and $OR_2 = e^{\beta_2}$), each ranging from 0.45 to 2.23, for each control arm distribution. The x and y axis are centered at 1, reflecting no treatment effect. The red vertical line at $OR_1 = 1$ represents the case where there is no difference in the odds of $Y = 1$ comparing the treatment groups, and the red horizontal line at $OR_2 = 1$ represents the case where there is no difference in the odds of $Y \leq 2$ comparing the two treatment groups. The 45 degree line represents the case where the proportional odds assumption holds (i.e. $OR_1 = OR_2$). Regardless of the control group distribution, the power to detect a treatment effect relies on both OR_1 and OR_2. When OR_1 and OR_2 are both 1, the cumulative logit model achieves the nominal 5% type I error rate, with the estimated power of 0.05, 0.06, 0.05 for the uniform, MISTIE II, and CLEAR II control distributions, respectively. In general, when fixing OR_1 (OR_2), the power to detect a significant treatment effect increases as $|OR_2|$ ($|OR_1|$) becomes large. Since the cumulative logit model simultaneously considers the effect of treatment at both $Y = 1$ and $Y = 2$, if at least one of OR_1 and OR_2 is far from 1, this model can detect a treatment effect, and thus results in high power. In fact, the estimates of the power are higher when OR_1 and OR_2 are not in the same direction with the center at 1 (e.g. $OR_1(OR_2) < 1$ and $OR_2(OR_1) > 1$), that is to say that the cumulative logit model can be best utilized when either one of the OR_1 and OR_2 is far away from 1, and the difference between OR_1 and OR_2 is large.

Figure 4.2: Heat maps of powers for the cumulative logit model



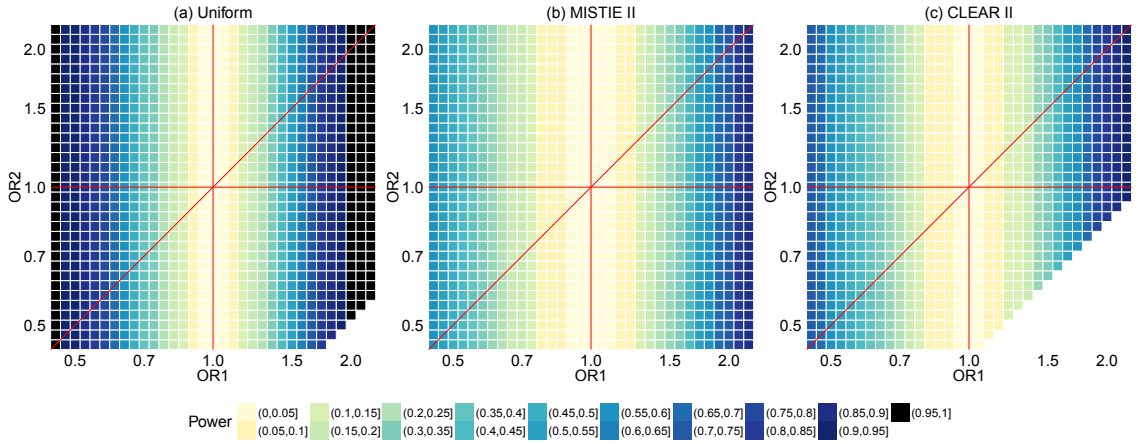
We observe different patterns in the power to detect a treatment effect when we change the control arm distribution. Regardless of the control arm distribution, the power is low to detect a treatment effect when OR_1 and OR_2 are both close to 1. For both the uniform and CLEAR II control arm distributions, we see that power increases more quickly as $|OR_1|$ ($|OR_2|$) increases and $|OR_2|$ ($|OR_1|$) decreases. For the MISTIE II trial, the power increases at a slower rate as $|OR_1|$ ($|OR_2|$) increases and $|OR_2|$ ($|OR_1|$) decreases. These features are attributable to the control arm distribution, and the size of the individual OR_1 and OR_2. We go into further details in Section 4.2.4.

4.2.2 Fixed dichotomy

Fixed dichotomy at $Y = 1$

When the outcomes are dichotomized to differentiate between $Y = 1$ vs. $Y > 1$, the power of Fisher's exact test is solely a function of OR_1 (Figure 4.3). This makes

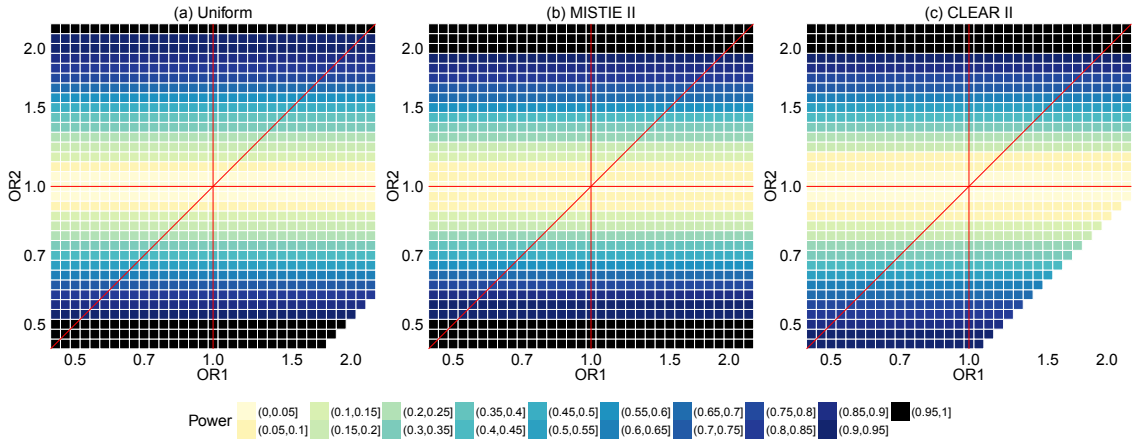
Figure 4.3: Heat maps of powers for the fixed dichotomy at Y=1



sense considering OR_1 compares the odds of $Y = 1$ and $Y > 1$ across the treatment groups and with a fixed dichotomy defined at $Y = 1$, OR_1 contains all the information to define the treatment effect, and as OR_1 moves away from 1, the power increases. At $OR_1 = 1$, the Fisher's exact test achieves the nominal 5% significance level, returning a power of 0.04, 0.04, 0.04 for the uniform, MISTIE II, and CLEAR II control distribution, respectively. When the assumed control distribution is MISTIE II or CLEAR II, the power to detect a treatment effect does not increase as quickly with an increase in $|OR_1|$ as it does when we assume the uniform control distribution. In the MISTIE II and CLEAR II trials, a smaller proportion of control patients have $Y = 1$ and for the same OR_1, the absolute difference in proportions of patients with $Y = 1$ comparing the treatment to control arm is smaller for these two trials compared to the uniform control distribution, hence the lower power.

Fixed dichotomy at $Y = 2$

Figure 4.4: Heat maps of powers for the fixed dichotomy at Y=2

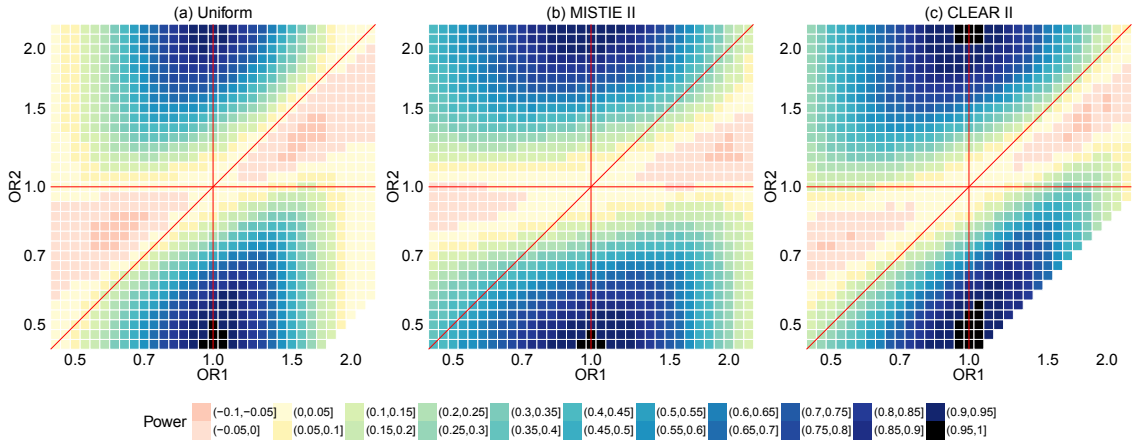


Similar intuition applies to the case when the outcomes are dichotomized at $Y = 2$, where power is determined only by the value of OR_2. For the fixed dichotomy at $Y = 2$, we ignore differences in the proportion of patients with $Y = 1$ in each treatment group, treatment comparisons are based only on the proportion of patients with $Y \leq 2$ in both treatment groups. In this case, when $OR_2 = 1$, the Fisher's exact test achieves the nominal 5% significance level, with the estimated power to be 0.04, 0.04, 0.04 for the uniform, MISTIE II, and CLEAR II control distribution, respectively. As expected, Figure 4.4 shows that the power to detect a treatment effect based on the fixed dichotomy defined by $Y = 2$ increases as OR_2 moves away from 1. When the uniform control distribution is assumed, the proportion of control patients in $Y \leq 2$ category is 67%, and this contributes to faster increases in power as OR_2 becomes smaller than 1 compared to when OR_2 becomes larger than 1. When

the MISTIE II control distribution is assumed, the power increase is almost identical regardless of direction of the treatment effect. This happens because the proportion of control patients in $Y \leq 2$ category is 55% in the MISTIE II trial. Whereas in the CLEAR II control distribution, the proportion of patients in $Y \leq 2$ category is still relatively small (27%), thus the power increase happens quicker as OR_2 becomes larger than 1.

4.2.3 Comparison between cumulative logit model and fixed dichotomy

Figure 4.5: Power comparison of the cumulative logit model and fixed dichotomy at Y=1

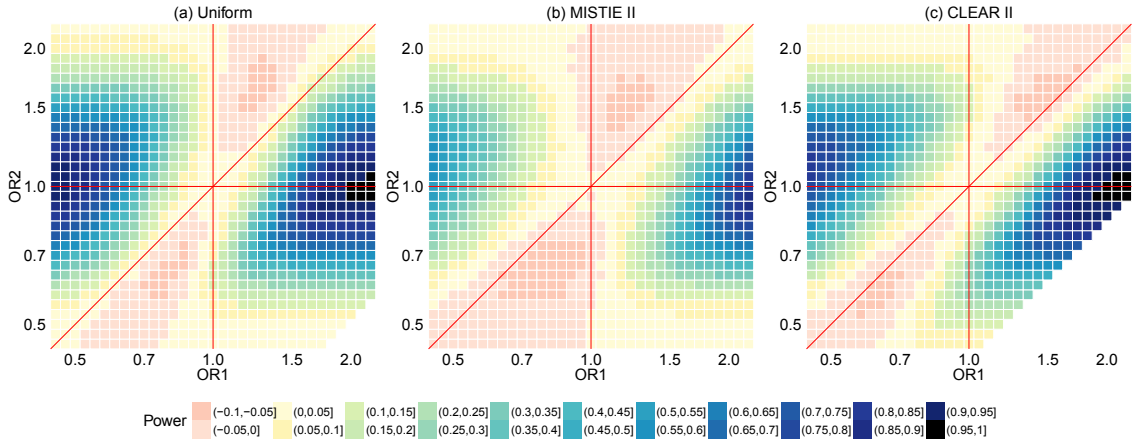


We compared the power of the test for treatment effect based on the cumulative logit model and the Fisher's exact test when the ordinal outcome is dichotomized at $Y = 1$ (Figure 4.5) and at $Y = 2$ (Figure 4.6). We subtracted the power to detect a treatment effect based on the Fisher's exact test from the power based on

the cumulative logit model, thus the red areas of the heat maps represent the cases where the Fisher's exact test outperforms the cumulative logit model in terms of power to detect a treatment effect. We note that overall, the cumulative logit model is not uniformly better than the fixed dichotomy at $Y = 1$.

Recall that the power to detect a treatment effect based on the fixed dichotomy at $Y = 1$ depends only on OR_1. In cases where OR_1 is close to 1 but OR_2 is further away from 1, then the cumulative logit model yields higher power to detect a treatment effect. In general, dichotomizing the ordinal outcome at $Y = 1$ provides a test for the treatment effect that has greater power than the test based on the cumulative logit model when $1 > OR_2 > OR_1$ and $1 < OR_2 < OR_1$.

Figure 4.6: Power comparison of the cumulative logit model and fixed dichotomy at Y=2



We also compared the power of the test for treatment effect based on the cumulative logit model and the Fisher's exact test when the ordinal outcome is dichotomized at $Y = 2$. For the fixed dichotomy at $Y = 2$, the power is a function of only OR_2 and

we see that when OR_2 is close to 1 but OR_1 is away from 1, there is an advantage to using the cumulative logit model. In addition, the fixed dichotomy generally yields larger power to detect a treatment effect compared to the cumulative logit model when $1 < OR_1 < OR_2$ and $1 > OR_1 > OR_2$.

4.2.4 Discussion of specific cases

In the prior sections, we demonstrated that the cumulative logit model (CLM) does not always provide higher power to detect treatment effects compared to the fixed dichotomy (FD) approaches. To further explore this finding, we have selected a few combinations of OR_1 and OR_2 values to demonstrate when and why the cumulative logit model or fixed dichotomy approaches are preferrable. For each selected set of OR_1 and OR_2, Table 4.1 displays the cumulative proportions of patients in each treatment arm as well as the estimated power based on the cumulative logit model and fixed dichotomies at $Y = 1$ and $Y = 2$ for each control arm distribution.

First, we considered two cases where the cumulative logit model had greater power to detect a treatment effect compared to the fixed dichotomy. When $OR_1 = 1.65$ and $OR_2 = 1$, the cumulative logit model has higher power to detect a treatment effect compared to the fixed dichotomy at $Y = 1$ when the control arm distribution is uniform or based on the CLEAR II trial. For the MISTIE II trial, the power is roughly the same for both the cumulative logit model and fixed dichotomy at $Y = 1$. In this case, there is no difference in the proportions of patients with $Y \leq 2$ between the treatment arms so that the differences in power are driven by the proportion of patients with $Y = 1$. The power of Fisher's exact test with

the dichotomy assigned at $Y = 1$ increases when two conditions are met: 1) when the risk difference, $Pr(Y \leq 1 \mid A = 1) - Pr(Y \leq 2 \mid A = 0)$ is larger, 2) the $Pr(Y \leq 1 \mid A = 0)$ is closer to 0.5. Thus the highest power of Fisher's exact test is achieved for the uniform control distribution since the risk difference, i.e. $Pr(Y \leq 1 \mid A = 1) - Pr(Y \leq 1 \mid A = 0)$, is the largest, and also $Pr(Y \leq 1 \mid A = 0$ is the closest to 0.5. The increase in power when the cumulative logit model is used, can be attributable to the fact that the test for a treatment effect in the cumulative logit model is looking at $Pr(Y = 1 \mid A = 1) - Pr(Y =\mid A = 0)$ but also at $Pr(Y = 2 \mid A = 1) - Pr(Y = 2 \mid A = 0)$. And with $OR_1 = 1.65$ and $OR_2 = 1$, the true risk differences are 12% and 11% respectively for the uniform control distribution, 5% and 5% for the MISTIE II control distribution, and 7% and 7% for the CLEAR II control distribution. Therefore, the cumulative logit model has a greater chance to detect at least one of these two risk differences being different from 0 than the fixed dichotomy has on detecting a single non-zero risk difference, with an exception for the MISTIE II control distribution, where the risk differences are small at both levels of outcome.

Alternatively, we considered the case where $OR_1 = 1$ and $OR_2 = 1.65$, where the cumulative logit model has higher power to detect a treatment effect compared to the fixed dichotomy at $Y = 2$ for the Uniform and CLEAR II control arm distribution. Similarly, we note that the Fisher's exact test when the dichotomy was assigned at $Y = 2$ gains more power when the risk difference, $Pr(Y \leq 2 \mid A = 1) - Pr(Y \leq 2 \mid A = 0)$ is larger, and the $Pr(Y \leq 2 \mid A = 0)$ is closer to 0.5. For instance,the highest power is achieved under the MISTIE II control arm distribution because

the risk difference is the largest with 12% and $Pr(Y \leq 2 \mid A = 0)$ is 0.55, being very close to 0.5. When we set $OR_1 = 1$ and $OR_2 = 1.65$, the cumulative logit model has higher power than the fixed dichotomy at $Y = 2$ with an exception for the MISTIE II control arm distribution, due to similar reasons discussed above; the cumulative logit model has increased power because it is looking for significant risk difference at least one value of Y, whereas the fixed dichotomy approach only looks for significant risk different at a specified cut-off of Y. In this case, the risk differences between the treatment groups at both $Y = 2$ and $Y = 3$ are consistent within each control arm distribution (10%, 12%, 11% for the Uniform, MISTIE II, and CLEAR II control distributions, respectively), incorporating risk differences at both levels provides more power to detect a treatment effect when the cumulative logit model is used, with an exception for the MISTIE II control arm distribution, where $Pr(Y \leq 2 \mid A = 0)$ is close to 0.5, yields higher power for the fixed dichotomous approach.

We also considered two scenarios where the fixed dichotomy was favored over the cumulative logit model. First consider $OR_1 = 1.65$ and $OR_2 = 1.28$. Since the $OR_1 > OR_2$ we expect to see higher power to detect a treatment effect based on the fixed dichotomy at $Y = 1$ relative to $Y = 2$. The fixed dichotomy at $Y = 1$ approach yields higher power than the cumulative logit model for the uniform and CLEAR II control arm distributions. This is attributable to the risk differences between the treatment groups across all 3 levels of outcomes; the risk differences are 12%, 6%, and 6% at $Y = 1$, $Y = 2$, $Y = 3$ under the uniform control distribution, and 7%, 1%, and 6% under the CLEAR II control distribution, where the largest risk difference

is found at $Y = 1$, resulting in the higher power when the fixed dichotomy at $Y = 1$ used to detect a treatment effect, compared to the cumulative logit model. On the contrary, the risk differences for the MISTIE II control arm distributions are 5%, 1%, and 6% at $Y = 1$, $Y = 2$, $Y = 3$, where the largest difference is found at $Y = 3$, thus using the cumulative logit model, which considers the risk difference at all 3 levels, yields higher power than the fixed dichotomy at $Y = 1$.

Lastly, we considered a case where the fixed dichotomy at $Y = 2$ has greater power than the cumulative logit model, $OR_1 = 1.28$ and $OR_2 = 1.65$. First we see that the power of Fisher's exact test to detect the treatment effect is higher when the dichotomy is assigned at $Y = 2$ than $Y = 1$ since $OR_2 > OR_1$. And when the cut-off for the fixed dichotomy is assigned at $Y = 2$, we see that the Fisher's exact test yields higher power than the cumulative logit model across all three control arm distributions for the similar reasons above. With $OR_1 = 1.28$ and $OR_2 = 1.65$, the risk differences are 6%, 4%, and 10% at $Y = 1$, $Y = 2$, $Y = 3$ under the uniform control arm distribution, 2%, 10%, and 12% under the MISTIE II control arm distribution, and 3%, 8%, and 11% under the CLEAR II control arm distribution. We see that the largest risk differences are at $Y = 3$, which are mathematically equivalent to the risk differences for $Y \leq 2$ when we have 3 outcome categories. Thus by dichotomizing the outcomes at $Y = 2$, the Fisher's exact test has an advantage in detecting the treatment effect since the largest risk difference is found at $Y \leq 2$.

Based on the four cases we present above, we noted that even with the same odds ratios, the proportion of treatment patients in each outcome category is different

based on the pre-specified control arm distribution. In addition, we highlighted several cases where the cumulative logit model was superior (inferior) to the fixed dichotomy approach in terms of power.

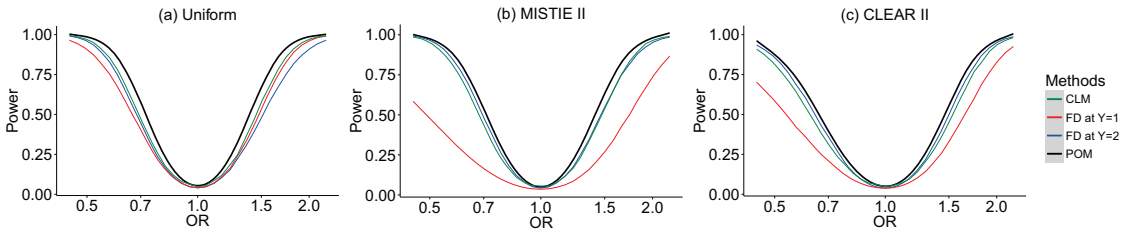
Table 4.1: Comparison of the $Pr(Y \leq k)$ for $k = 1, 2$ and the power to detect a treatment effect assuming assuming different odds ratios

Control arm distribution	Pr(Y≤ 1)		Pr(Y≤2)		Power to detect the treatment effect		
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	CLM	FD at Y=1	FD at Y=2
$OR_1 = 1.65$ and $OR_2 = 1$							
Uniform	0.33	0.45	0.67	0.67	0.86	0.75	0.04
MISTIE II	0.11	0.16	0.55	0.55	0.419	0.421	0.04
CLEAR II	0.14	0.21	0.27	0.27	0.81	0.51	0.04
$OR_1 = 1$ and $OR_2 = 1.65$							
Uniform	0.33	0.33	0.67	0.77	0.70	0.04	0.67
MISTIE II	0.11	0.11	0.55	0.67	0.72	0.04	0.74
CLEAR II	0.14	0.14	0.27	0.38	0.83	0.04	0.71
$OR_1 = 1.65$ and $OR_2 = 1.28$							
Uniform	0.33	0.45	0.67	0.72	0.68	0.75	0.22
MISTIE II	0.11	0.16	0.55	0.61	0.43	0.42	0.25
CLEAR II	0.14	0.21	0.27	0.33	0.46	0.51	0.22
$OR_1 = 1.28$ and $OR_2 = 1.65$							
Uniform	0.33	0.39	0.67	0.77	0.60	0.24	0.67
MISTIE II	0.11	0.13	0.55	0.67	0.68	0.12	0.74
CLEAR II	0.14	0.17	0.27	0.38	0.69	0.14	0.71

4.2.5 When the proportional odds assumption holds

We also considered the special case where the proportional odds assumption holds ($OR_1 = OR_2$) and compared the power to detect a treatment effect using the proportional odds model (a 1 degree of freedom test for OR) with the Fisher's exact test for the fixed dichotomy at $Y = 1$ or $Y = 2$, also with the cumulative logit model (a 2 degree of freedom test for OR_1 and OR_2). Figure 4.7 displays the estimated powers

Figure 4.7: Power comparison when the proportional odds assumption holds



centered at $OR = 1$, based on different statistical methods across the control arm distributions. Given our sample size of $n = 500$ patients, for each of the assumed control group distributions, the proportional odds model outperforms the fixed dichotomous approach in terms of power to detect a treatment effect. Comparing the two fixed dichotomy approaches, in general, the power is higher when the dichotomy is assigned at $Y = 2$ relative to $Y = 1$. However, for the uniform control distribution, the fixed dichotomy at $Y = 1$ yields greater power than the fixed dichotomy at $Y = 2$ when the $OR > 1$. Also, the cumulative logit model yields lower power than the proportional odds model due to the extra 1 degree of freedom in the cumulative logit model.

We created an example table to understand the results more numerically. Table 4.2 presents cumulative proportions in each treatment group across the three control arm distributions as well as the power estimates when we assume $OR = 1.65$. By incorporating the treatment effect happening at both levels in the outcomes, but not paying penalty for an extra degree of freedom as in the cumulative logit model, the

proportional odds model (POM) achieves the highest power to detect the treatment effect when the proportional odds assumption is true.

Table 4.2: Comparison of the $Pr(Y \leq k)$ for $k = 1, 2$ and the power to detect a treatment effect assuming $OR = 1.65$

Control arm distribution	Pr(Y≤1)		Pr(Y≤2)		Power to detect the treatment effect			
	$A = 0$	$A = 1$	$A = 0$	$A = 1$	POM	FD at Y=1	FD at Y=2	CLM
Uniform	0.33	0.45	0.67	0.77	0.86	0.75	0.67	0.78
MISTIE II	0.11	0.16	0.55	0.67	0.83	0.42	0.74	0.75
CLEAR II	0.14	0.21	0.27	0.38	0.76	0.51	0.71	0.66

Chapter 5

Conclusions

We sought to identify characteristics of the distributions of the control and treatment group that provide higher power to detect a treatment effect when applying a fixed dichotomy to an ordered categorical outcome. The simulation study considered an ordered categorical outcome with three levels and three different control group distributions.

When choosing between the ordinal approach vs. fixed dichotomous approach, it is important to understand where the treatment effect is among the levels of outcomes. When it is reasonable to assume proportional odds, using the proportional odds model with the ordinal outcome yields higher power to detect a treatment effect compared to a fixed dichotomous approach or the cumulative logit model.

More careful consideration is required when the proportional odds assumption does not hold. In such cases, we found that the fixed dichotomous approach is superior than the ordinal approach when the dichotomy of the outcome is assigned at the level c, where the treatment effect maximizes the risk difference between $Pr(Y \leq c \mid A = 1)$ and $Pr(Y \leq c \mid A = 0)$. In addition to the risk difference at the

level c, the Fisher's exact test gains power to detect a treatment as $Pr(Y \leq c \mid A = 0)$ being closer to 0.5.

The cumulative logit model yeilds higher power copmared to the proportional odds model and fixed dichotomy approach when the treatment group comparisons occurring at each level of the outcome have different directions (i.e. control is favored at one cut-point but then treatment is favored at the next cut-point).

We have considered an ordinal outcome with three levels; however, in practice, many ordinal scales used in practice may have more then 3 levels. This work needs to be extended to the general case. In addition, in practice, the cumulative logit model is seldom implemented. Instead, the proportional odds model is more frequently adapted even when the proportional odds assumption is not valid. How to compare the power to detect a treatment effect when the wrong model is implemented is a further area of interest.

References

[1] Douglas G. Atlman. *Practical Statistics for Medical Research*. Chapman and Hall, 1991.

[2] Philip M W Bath, Kennedy R. Lees, Peter D. Schellinger, Hernan Altman, Martin Bland, Cheryl Hogg, George Howard, and Jeffrey L. Saver. Statistical analysis of the primary outcome in acute stroke trials. *Stroke*, 43(4):1171–1178, 2012.

[3] Sidney Siegel and Jr. N.John Castellan. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, Inc., second edition, 1988.

[4] F Wilcoxon. Individual comparisons of grouped data by ranking methods. *Journal of economic entomology*, 39(6):269, 1945.

[5] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947.

[6] Rollin Brant. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 46(4):1171–1178, 1990.

[7] Joseph L.Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, Inc., third edition, 2003.

[8] Daniel Hanley; NINDS, Inc. Genentech, and Emissary International LLC. Minimally Invasive Surgery and rtPA for Intracerebral Hemorrhage Evacuation (MISTIE), 2005.

[9] Ruth Bonita and Robert Beaglehole. Modification of rankin scale: Recovery of motor function after stroke. *Stroke*, 19(12):1497–1500, 1988.

[10] Johns Hopkins University; FDA Office of Orphan Products Development and Inc. Genetech. Clinical Trial on Treatment of Intraventricular Hemorrhage (CLEAR IVH), 2007.

Vita

So Yung Choi was born in Busan, Republic of Korea, on February 18, 1990. She received a Bachelor of Arts in Economics from the University of Hawaii at Manoa with a certificate in Linguistics in 2012. She began work on her Master of Science in Biostatistics and a certificate in Risk Sciences and Public Policy in 2014 at the Johns Hopkins Bloomberg School of Public Health.